Centurion: Building a 50-Gigaflop Computer from Commercial, Off-the-Shelf Components

Steve J. Chapin
Department of Computer Science
Thornton Hall
University of Virginia
Charlottesville, VA 22903
phone: (804) 982-2220 fax: (804) 982-2214 email: chapin@cs.virginia.edu

Award #: N00014-98-1-0454

http://legion.virginia.edu/centurion/Centurion.html

LONG-TERM GOAL

Our long-term goal is to investigate the potential of multicomputers constructed of commercial, off-the-shelf components as a replacement for conventional parallel processors in high-performance computing. By commercial, off-the-shelf components, we mean commodity workstations and personal computers connected by readily available networking fabric, e.g. DEC Alpha and Intel PC boxes connected by gigabit ethernet. To understand the impact of clusters in this area, we are developing several layers of software to support traditional supercomputing applications such as climate and ocean modeling.

OBJECTIVES

We hope to demonstrate the efficacy of constructing medium-scale, clustered multicomputers for a fraction of the cost of traditional supercomputers. We are building a clustered system called Centurion. Using Centurion, we will demonstrate that we can outperform a cluster of C90 machines, and at a fraction of the cost. This research can point the way to developing more effective, lower-cost computational engines for the Department of Defense.

APPROACH

While the original proposal was to build a 50-Gigaflop machine, the resultant machine will have more than a 200 Gigaflop peak rate. We are building this multicomputer using off-the-shelf commodity processors based on both DEC Alpha and Intel commodity processors. One third of the machine is connected in a 2D torus using 1.28 Gb/s Myrinet switches; all processors are interconnected by fast ethernet switches and gigabit ethernet hubs. This machine will be used to solve computationally challenging science and engineering problems and as a testbed for systems-oriented computer science research. The heterogeneous nature of the machine, both in terms of processor architecture and networking fabric, gives us a rich environment for determining the best operational aspects of our application suite.

Our application suite consists of two ocean simulations, a directed vapor deposition code, a DNA sequence comparison code, a polyatomic system simulator, a biomolecular simulation, a macromolecular dynamics and mechanics code, and a system for performing molecular orbital calculations. For details, see http://legion.virginia.edu/centurion/Applications.html.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **1998**	2. REPORT TYPE	3. DATES COVERED **00-00-1998 to 00-00-1998**

4. TITLE AND SUBTITLE **Centurion: Building a 50-Gigaflop Computer from Commercial, Off-the-Shelf Components**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Virginia,Department of Computer Science,Charlottesville,VA,22903**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

See also ADM002252.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**5**	

WORK COMPLETED

We have begun acquisition of the required equipment, having issued requests for bids on 66 533-MHz Alpha machines and 66 350 MHz dual-processor Intel Pentium II machines. We have received approximately 3 dozen responses to these requests, and are in the process of reconciling the bids with the requirements to determine the vendor which will be awarded the contract. We are also in the process of acquiring the necessary Ethernet switches to interconnect the new machines and join them to our existing cluster.

In the meantime, we have ported the Navy Layered Ocean Model (NLOM) to our existing cluster of 64 DEC Alpha machines, and have obtained the Miami Isopycnic Coordinate Ocean Model (MICOM) software for porting to Centurion. We have also ported an axially-symmetric direct simulation Monte Carlo code to the machine.

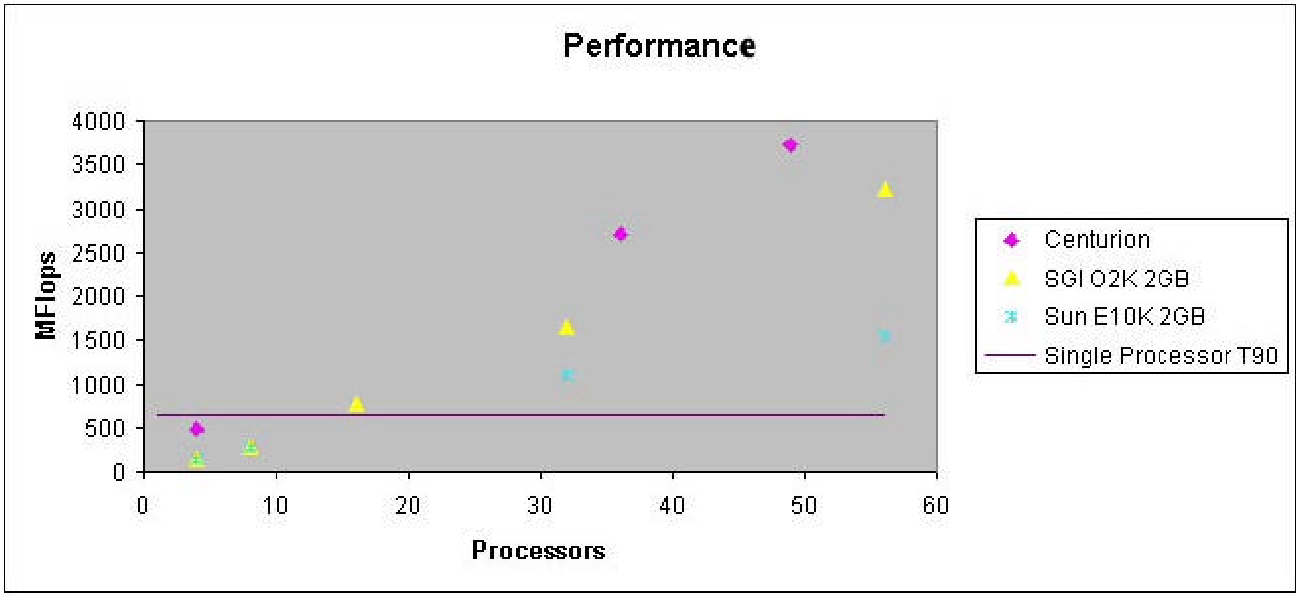
RESULTS

It is too early in the acquisition cycle for Centurion for us to have meaningful results on the full cluster. Our 1999 Annual Report will contain analysis of the software performance achieved on the new cluster.

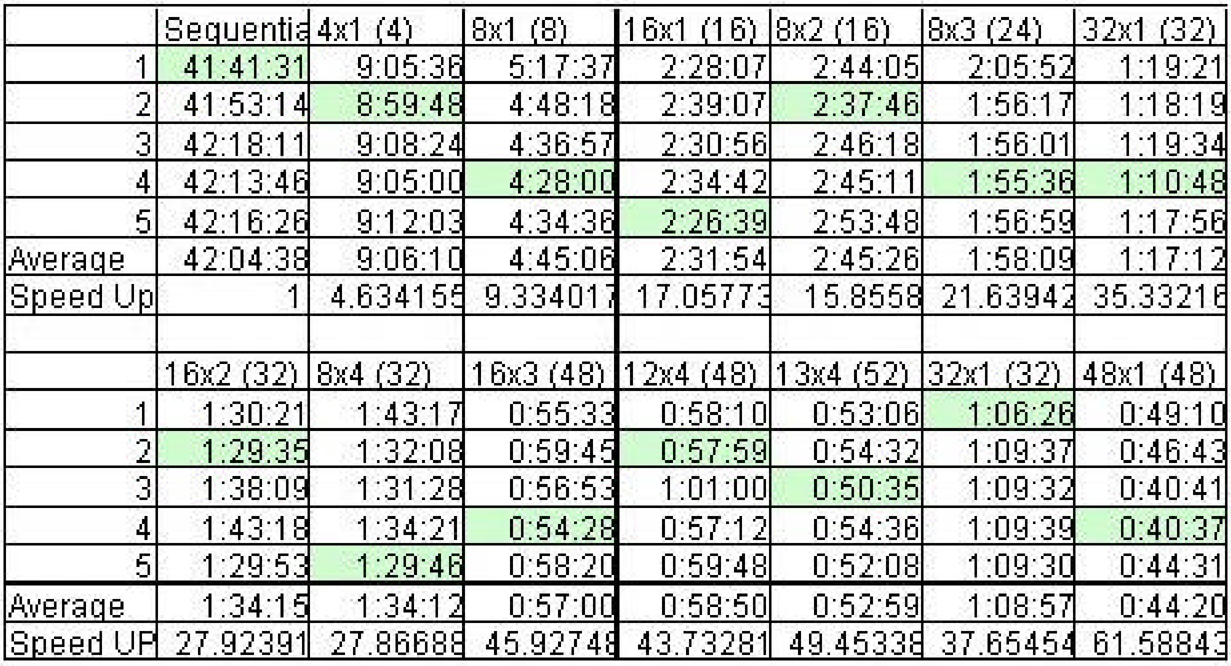
We can report results from the NLOM code on the existing Centurion cluster. In this case, we started with an unoptomized version of the code, and benchmarked on 4 processors within Centurion. The following table lists the various optimizations made and their effects on performance.

Description	Performance
Out of the box	~70 Mflops on 4 nodes
Cache block 2 main loops, split communication	~175 Mflops
Inter-procedural analysis, eliminate 2 matrix copies	~195 Mflops
Eliminate communcation associated with extra copies	~235 Mflops
Loop jam	~220 Mflops
Cache block jammed loop	~355 Mflops
Fix Legion-MPI bug	~480 Mflops

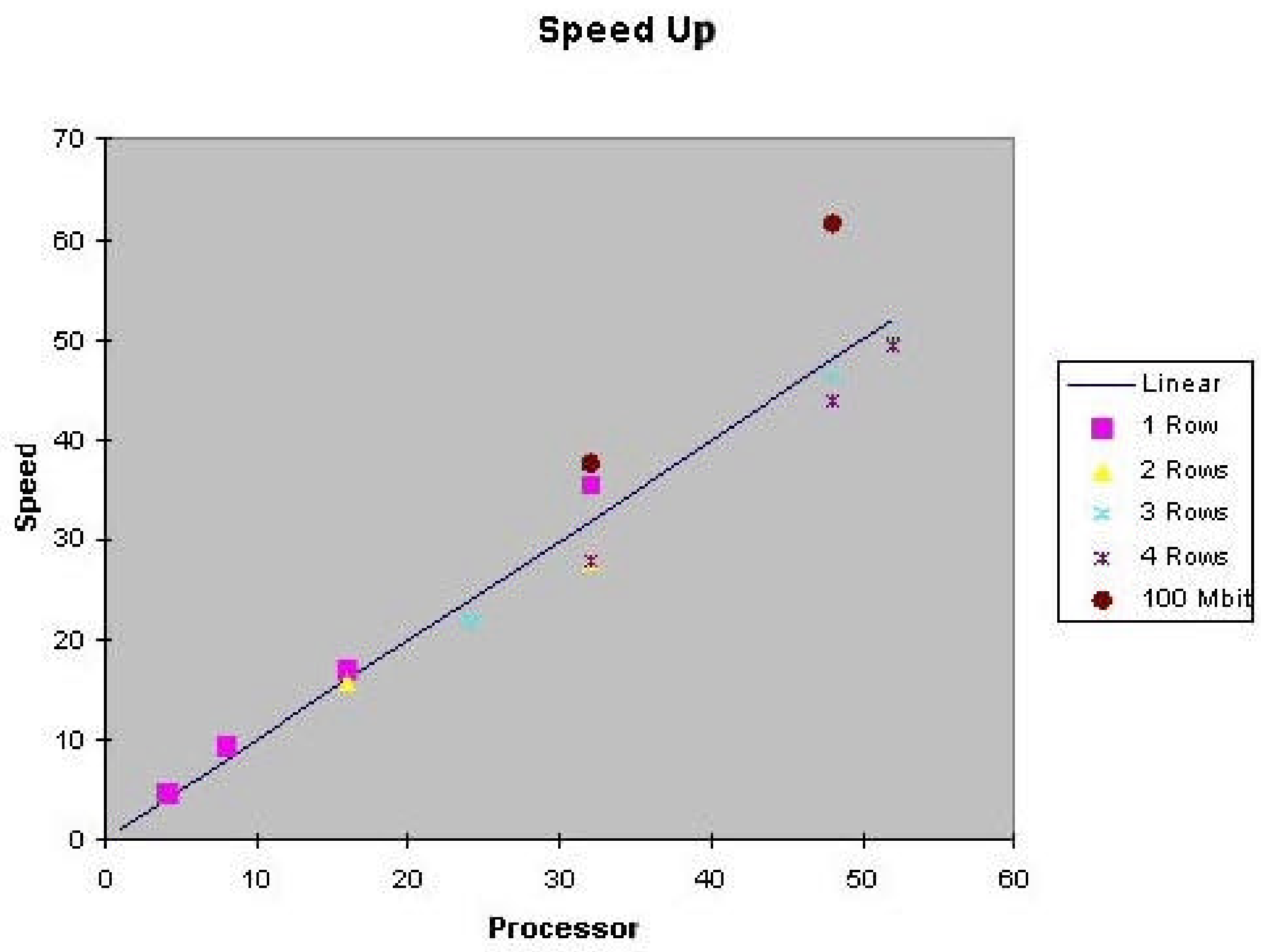
In addition, we ran the ocean model on 49 processors. The results of this run are depicted in the following graph. It should be noted that we achieved 3.7GF with $150,000 worth of equipment.



We also measured performance for an axially-symmetric steady flow materials simulation. This program simulates the flow of a gas jet during chip manufacture; the intent is to improve fabrication yield. The original code took more than one week to execute on a single RS/6000. On 52 nodes of the Centurion cluster, it took 41 minutes after parallelization. We did observe superlinear speedup because of cache effects. The speedup results appear in the following table and graph; full details are in [1].



	Sequentia	4x1 (4)	8x1 (8)	16x1 (16)	8x2 (16)	8x3 (24)	32x1 (32)
1	41:41:31	9:05:36	5:17:37	2:28:07	2:44:05	2:05:52	1:19:21
2	41:53:14	8:59:48	4:48:18	2:39:07	2:37:46	1:56:17	1:18:19
3	42:18:11	9:08:24	4:36:57	2:30:56	2:46:18	1:56:01	1:19:34
4	42:13:46	9:05:00	4:28:00	2:34:42	2:45:11	1:55:36	1:10:48
5	42:16:26	9:12:03	4:34:36	2:26:39	2:53:48	1:56:59	1:17:56
Average	42:04:38	9:06:10	4:45:06	2:31:54	2:45:26	1:58:09	1:17:12
Speed Up	1	4.634155	9.334017	17.05773	15.8558	21.63942	35.33216
	16x2 (32)	8x4 (32)	16x3 (48)	12x4 (48)	13x4 (52)	32x1 (32)	48x1 (48)
1	1:30:21	1:43:17	0:55:33	0:58:10	0:53:06	1:06:26	0:49:10
2	1:29:35	1:32:08	0:59:45	0:57:59	0:54:32	1:09:37	0:46:43
3	1:38:09	1:31:28	0:56:53	1:01:00	0:50:35	1:09:32	0:40:41
4	1:43:18	1:34:21	0:54:28	0:57:12	0:54:36	1:09:39	0:40:37
5	1:29:53	1:29:46	0:58:20	0:59:48	0:52:08	1:09:30	0:44:31
Average	1:34:15	1:34:12	0:57:00	0:58:50	0:52:59	1:08:57	0:44:20
Speed UP	27.92391	27.86688	45.92748	43.73281	49.45338	37.65454	61.58843



IMPACT/APPLICATION

This DURIP award will more than triple the available computing power in our current Centurion cluster. This will allow us to make real-world computational and bang-for-the-buck comparisions between our cluster, obtained for a total investment of less than $1,000,000, to that of the Cray machines at the NaVO MSRC in Stennis, MS on the same applications.

TRANSITIONS

It is too early in the project for anyone to have adopted our methods. When we have installed the full system and performed experimental analysis, we will present our results to the various DoD MSRCs and Research Labs. We hope that, as a result of this work, cluster computing will become a mainstream computing resource for the DoD.

RELATED PROJECTS

1 – We are developing the Legion run-time system for use on the Centurion cluster. Legion will provide automated binary management, an object-oriented computing paradigm (but also supporting

legacy code in Fortran), security, resource management, and multi-language/environment support, including MPI, PVM, Fortran, C++, and Java.

2 – Under the HOSS project, I am developing operating system mechanisms to support high-performance computing on clusters. My current focus is on flexible address-space (distributed shared memory) management, in collaboration with Mike Hereoux of Sandia National Labs and David Bader of the University of New Mexico.

REFERENCES

[1] Beekwilder, N. and Grimshaw, A. *Parallelization of an Axially Symmetric Steady Flow Program*, Computer Science Technical Report #CS-98-10, University of Virginia, May, 1998. ftp://ftp.cs.virginia.edu/pub/techreports/CS-98-10.ps.Z.

[2] Lindahl, G., Chapin, S.J., Beekwilder, N., and Grimshaw, A. Experiences with Legion on the Centurion Cluster, Computer Science Technical Report #CS-98-27, University of Virginia, August, 1998. ftp://ftp.cs.virginia.edu/pub/techreports/CS-98-27.ps.Z.

PUBLICATIONS

None.